Forensic Science International: Digital Investigation 53 (2025) 301885

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

More on digital evidence exceptionalism: Critique of the argument-based method for evaluative opinions

Alex Biedermann [a,*], Kyriakos N. Kotsoglou [b]

[a] University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, 1015 Lausanne–Dorigny, Switzerland
[b] University of Northumbria, School of Law, Newcastle Upon Tyne, NE1 8ST, United Kingdom

A B S T R A C T

This paper critically analyses and discusses the "Argument-Based Method for Evaluative Opinions" (ABMEO) recently proposed by Sunde and Franqueira in a paper published in *Forensic Science International: Digital Investigation* (Sunde and Franqueira, 2023). According to its developers, this novel method allows one to produce evaluative opinions in criminal proceedings by constructing arguments. The method is said to incorporate concepts from argumentation and probability theory, while ensuring adherence to accepted principles of evaluative reporting, in particular the ENFSI Guideline for Evaluative Reporting in Forensic Science. While this sounds promising, our analysis of the ABMEO, as well as Sunde and Franqueira's account of a number of evidence-related concepts such as probative value (and its assessment), credibility, relevance, normativity, and probability, among others, reveals a number of fundamental problems that are indicative of *digital evidence exceptionalism*; i.e. the idea that digital forensic science can somehow exempt itself from adhering to methodologically and scientifically rigorous evidence evaluation procedures. In this paper we explain why the ABMEO cannot and should not be considered as an appropriate complement, supplement or replacement for the existing reference framework for evaluative reporting in forensic science. In particular, we argue that the ABMEO is internally contradictory and tends to undermine the substantial progress made over the past two decades in the development and implementation of principles for the evaluative reporting of forensic science evidence.

> To the statement "I feel in my hand that the water is three feet under the ground" we should like to answer: "I don't know what this *means*".
>
> [Ludwig Wittgenstein]

1. Introduction

As the use of forensic science in legal proceedings increasingly involves digital evidence, it is vital to ensure that its probative value is assessed using logically sound principles. Principles for evaluating items of forensic evidence, regardless of their nature (e.g., biological traces, marks, impressions etc.), have been well established for some time, first in the scientific literature (e.g. Aitken et al., 2020; Evett and Weir, 1998; Robertson et al., 2016) and then in a number of high-profile guidelines and recommendations for practitioners (e.g. Aitken et al., 2010; Gill et al., 2018; Nic Daéid et al., 2020; Willis et al., 2015). The application of these principles in practice is challenging, not only in some of the most scientifically advanced fields, such as forensic genetics (see Hicks et al. (2022) for a review), but also in some of the oldest forensic disciplines, such as friction ridge analysis (e.g. Champod et al., 2016; Swofford et al., 2021). The emerging field of digital forensic science (hereafter DFS) is no exception. Consumers of digital forensic evidence, particularly the judiciary, would therefore be well advised to keep a critical eye on how the discipline deals with these challenges.

If not forensic practice as such, one might at least expect the peer-reviewed literature on DFS to demonstrate a high level of understanding of, and adherence to, sound evaluation principles. Unfortunately, the DFS literature is not a safe haven in this regard. It is a fertile ground for misconceptions, ad hoc theories and ideas that pose impediments to the implementation of coherent evaluative reporting schemes. For example, instead of a scientific approach to dealing with uncertainty in the evaluation of digital evidence, through probability theory, alternative concepts such as "Digital Evidence Certainty Descriptors" (DECDs) (Horsman, 2020) have been proposed. DECDs are openly non-probabilistic and amount to what we have elsewhere criticised as *digital evidence exceptionalism* (Biedermann and Kotsoglou, 2020),[1] i.e. the idea that the discipline of DFS can somehow exempt itself from adhering to methodologically and scientifically rigorous evidence evaluation procedures because of its supposedly special status relative to other mainstream forensic science disciplines that adhere to basic methodological principles.

* Corresponding author.
E-mail address: alex.biedermann@unil.ch (A. Biedermann).

[1] See also Edmond and Martire (2018), Murphy (2009) and Saks (2007) for further examples of the use of this term in the forensic science and legal literature.

https://doi.org/10.1016/j.fsidi.2025.301885
Received 24 April 2024; Received in revised form 10 October 2024; Accepted 9 February 2025

Table 1
Abbreviations used in this paper.

ABMEO	Argument-based Method for Evaluative Opinions (Sunde and Franqueira, 2023)
DFS	Digital forensic science
ENFSI Guideline	ENFSI Guideline for Evaluative Reporting in Forensic Science (Willis et al., 2015)
LFI	Logical framework for interpretation (e.g. Aitken et al., 2010, 2020; Evett, 2015; Robertson et al., 2016)
PV	Probative value
SF	Sunde and Franqueira (Sunde and Franqueira, 2023)

More recently, another proposal for producing evaluative conclusions has found its way into the DFS literature. Sunde and Franqueira (SF) propose the introduction of what they call "a novel method for producing argument-based evaluative opinions in the context of criminal investigation" (Sunde and Franqueira, 2023, p. 1), by drawing "on general concepts from argumentation theory, combined with key principles and concepts from probabilistic and narrative/scenario approaches to develop arguments and analyse evidence" [p. 1]. While this sounds promising, in particular the claimed reliance on "principles and concepts from probabilistic [...] approaches" (Sunde and Franqueira, 2023, p. 1), a closer examination of their methodology reveals a number of conceptual problems that run counter to the reference framework for evaluative reporting as given, for example, by the ENFSI Guideline for Evaluative Reporting (Willis et al., 2015) (hereafter referred to as ENFSI Guideline). The latter focuses on a probabilistic metric for quantifying the strength of evidence, also known as the logical framework for interpretation (e.g. Evett, 2015) (LFI).

The purpose of this paper is to outline and discuss what we consider to be the major problems associated with SF's Argument-Based Method for Evaluative Opinions (hereafter ABMEO), and to explain why it cannot and should not be considered an appropriate complement, supplement or substitute for the reference framework for evaluative reporting in forensic science. Specifically, we argue that the ABMEO is internally contradictory and undermines the progress made over the past two decades in developing and implementing a probability-based framework for evaluative reporting.

This paper is structured as follows. Section 2 provides a brief introduction to SF's proposed "Argument-Based Method for Evaluative Opinions" (ABMEO). Section 3 critically examines and discusses the ABMEO, as well as SF's presentation of evidence-related concepts such as probative value (and its assessment), credibility, relevance, normativity, probability, bias, and selected jurisdictional aspects of England and Wales, in particular the admissibility of opinion evidence and the standard of proof. Table 1 provides a summary of the abbreviations used in this article.

2. The "Argument-Based Method for Evaluative Opinions" (ABMEO) proposed by Sunde and Franqueira

The method proposed by SF, called the "Argument-Based Method for Evaluating Opinions" (ABMEO), involves several steps, starting with the formulation of competing propositions representing the views of the prosecution and the defence. Subsequent steps involve specifying the components of the argument, consisting of the *evidence* and what SF call the *grounds* and *inferential leaps*. An argument is then evaluated using the notions of *credibility* and *relevance*. The final steps involve identifying attacks on the components of the argument, reassessing the arguments and formulating an opinion.

Two aspects of the ABMEO are worth noting. One is that the ABMEO emphasises the construction and analysis of individual arguments for each hypothesis. These arguments are then compared in the final step of the procedure. The second aspect concerns a sub-procedure for assessing credibility (of "evidence" and "ground") and relevance, followed by a combination of credibility and relevance assessments to derive an assessment of an argument as a whole, using a concept called Argument Matrix.

SF claim that the ABMEO "complies with the ENFSI principles" (Sunde and Franqueira, 2023, p. 4), in particular the "ENFSI requirements of balance, logic, robustness and transparency" (Sunde and Franqueira, 2023, p. 2). They also state that the ABMEO incorporates "key principles and concepts from probability" (Sunde and Franqueira, 2023, p. 9). As we will show in the next section, these claims are unfounded.

3. Critique of the ABMEO

In this section we critically review the claimed properties of the ABMEO, using the logical framework for interpretation (LFI) and guidance documents based on this framework (e.g. Willis et al., 2015) as a reference point. We also review the way in which SF present selected concepts of legal evidence and proof analysis in relation to the ABMEO. We structure our analysis as a series of discussion points. The order in which we present the discussion points does not imply any order of importance.

3.1. The notion of probative value

Throughout their paper, SF use the term *probative value* in relation to multiple targets, which prevents this term from having a clear meaning. Specifically, the term probative value is used in relation to the following three targets:

- *Evidence*: For example, the abstract mentions "an Argument Matrix for a holistic determination of the probative value of the evidence" (Sunde and Franqueira, 2023, p. 1). In addition, Section 1.1. of SF's article mentions the ABMEO "as a method for the structured evaluation of the probative value of evidence" (Sunde and Franqueira, 2023, p. 2).
- *Arguments*: In Section 2.2. of their article, SF mention "the *Argument Matrix* for a holistic determination of the probative value of the argument" (Sunde and Franqueira, 2023, p. 4, emphasis as in original).[2] Overall, SF use the term probative value mostly in describing the various steps that make up the ABMEO. For example, Step 3 of the ABMEO is described as "Assess the probative value of arguments" (Sunde and Franqueira, 2023, p. 5) and Step 5 as "Re-assess the probative value of arguments" (Sunde and Franqueira, 2023, p. 5).
- *Hypotheses*: In their illustration of the ABMEO using two examples, SF repeatedly speak of the "probative value of Hypothesis(a)" and the "probative value of Hypothesis(b)" (Sunde and Franqueira, 2023, p. 6, 7 and 8). These expressions are found in the presentation of the Steps 3 and 5 of the ABMEO, respectively. This is strange and confusing, however, because the titles of these steps use the term probative value in relation to *arguments*, not hypotheses.

This level of confusion is surprising and a cause of concern. Clearly, *only* the evidence (i.e. the forensic findings) can have probative value, which, however, may be nil. Stated otherwise, the evidence is the target to which we seek to assign probative value in relation to competing propositions through the construction of an argument, not the other way

[2] As an aside, note that this in direct contradiction with the abstract where SF (Sunde and Franqueira, 2023) mention the Argument Matrix in relation to the *evidence*, not arguments.

round. The hypotheses, we cannot stress this enough, are the *probanda*, and therefore by definition *cannot* have probative value themselves. The same holds for arguments.

The ENFSI Guideline, to which SF claim to adhere, is also sufficiently clear on this point. It states that an evaluative report "provides, ultimately, an assessment of the strength to be attached to the *findings* in the context of the alleged circumstances"[3] (Willis et al., 2015, p. 5, emphasis added) and that "[t]he forensic practitioner seeks to evaluate *findings* with respect to particular competing propositions set by the specific case circumstances or as indicated by the mandating authority" (Willis et al., 2015, p. 6, emphasis added). At no point does the ENFSI Guideline suggest that the term "probative value" can, let alone should, be applied to either propositions or arguments.

This conceptual clarity stems from the fact that the ENFSI Guideline defines the strength of support of the findings exclusively in terms of the likelihood ratio: "the expression of the extent to which the observations (i.e. findings) support one of the two competing propositions. The extent of the support is expressed [...] in terms of the magnitude of the likelihood ratio." (Willis et al., 2015, p. 27) The ENFSI Guideline therefore strictly adheres to the definition of the likelihood ratio, which focuses on the findings, given the propositions (e.g. Aitken et al., 2020). A statement such as "probative value of the hypotheses", as used by SF, has no justification in this context.

In summary, thus, we see that on the mere definitional question of a seemingly simple term, i.e. "probative value", SF's account does *not* comply with the ENFSI Guideline, contrary to what they claim.

3.2. Assessing the probative value

In the previous section, we saw that SF do not use the fundamental term *probative value* in accordance with the logical framework of interpretation (LFI). They also diverge on the question of how probative value should be assessed. SF's approach, the ABMEO, recommends "taking one perspective at a time" (Sunde and Franqueira, 2023, p. 9), i.e. assessing probative value separately and individually with respect to each proposition, leading to the assertion of *multiple* probative values for the *same* set of evidence. Thus, SF's claim here is that probative value can be assessed in relation to propositions in isolation.

None of this is supported by the LFI emphasised in the ENFSI Guideline. In the LFI, which adheres to the logic of probability theory, probative value *cannot* be asserted in relation to propositions in isolation. Instead, probative value arises only *after* the probability of the evidence has been articulated and compared under at least two competing propositions. Moreover, under a given set of assumptions, there is only *one* probative value, not multiple ones, for a given piece (or pieces) of evidence under consideration.

We therefore find that the ABMEO does not comply with the ENFSI Guideline, not only with respect to the definition of probative value (see our discussion in Section 3.1), but also with respect to the question of how to *assess* probative value. SF's claim that the ABMEO "complies with the ENSI principles for evaluative opinions" (Sunde and Franqueira, 2023, p. 4) because of the requirement to define a pair of competing propositions as part of Step 1 of the ABMEO is therefore short-sighted. For a logically sound evaluation of evidence, it is not enough simply to define a pair of propositions. Equally important is how these propositions are subsequently treated during evaluative reasoning. This is discussed further in the next section.

3.3. Transposing the conditional

The convoluted use of terms such as probative value for multiple targets such as evidence and propositions leads SF to make several statements that amount to transposing the conditional – regrettably. For example, SF write that "[t]he suggested method [ABMEO] involves the assessment of at least two opposing hypotheses (propositions)" (Sunde and Franqueira, 2023, p. 4). In addition, they write: "The relevance of an item of evidence indicates how strongly does it [sic] support the hypothesis in question, and is assessed by asking: *Assuming that the evidence is true, what is the probability that the hypothesis is true?*" (Sunde and Franqueira, 2023, p. 4, emphasis as in original).[4]

This is clearly not in accordance with the ENFSI Guideline. Evaluative reporting, according to the ENFSI Guideline, under no circumstances amounts to an assessment of propositions, let alone probabilities for propositions. Forensic scientists ought to evaluate findings given at least a pair of contrasting propositions, but refrain from assessing propositions themselves. Recall that the third principle of interpretation states that "[t]he role of the forensic scientist is to consider the probability of the evidence given the propositions" (Evett et al. (2000, p. 234) and Willis et al. (2015, p. 23)), not the reverse. The importance of this cannot be overstated, as it has been described as "the most important lesson for the logic of evaluative forensic science" (Evett, 2009, p. 159). The statement "*Assuming that the evidence is true, what is the probability that the hypothesis is true*" (Sunde and Franqueira, 2023, p. 4, emphasis as in original), formally written $\Pr(H \mid E)$, is a textbook example of the transposition of conditional probabilities.

3.4. On credibility, relevance and the Argument Matrix

We now turn our attention to a sub-procedure of the ABMEO that deals with "[e]valuating the probative value of arguments" (Sunde and Franqueira, 2023, p. 3), based on the assessment of two criteria, credibility and relevance, and their combination through what SF call the "*Argument Matrix* for a holistic determination of the probative value of the argument" (Sunde and Franqueira, 2023, p. 3, emphasis as in original). We will first discuss the two concepts of credibility and relevance, and then their combination through the Argument Matrix. In order to assess these concepts against fundamental standards such as logical structure, i.e. to demonstrate the concepts' lack thereof, we will introduce and use some minimal formal notation. Although the ABMEO is non-symbolic and non-numerical, we consider that formal notation is necessary to ensure precision in our analysis.

Let us start with the concept of credibility. SF state that credibility:

- "is the extent to which the ground and the associated evidence may be believed to be true" (Sunde and Franqueira, 2023, p. 3);
- "is assessed without considering the hypothesis in question, and by asking: *What is the probability that the ground is true?*" (Sunde and Franqueira, 2023, p. 3, emphasis as in original);
- "is an expression of subjective and empirically founded non-numerical probability, i.e. an expression of the evaluator's belief based on the empirical findings, as opposed to aleatory or statistically determined numerical probability" (Sunde and Franqueira, 2023, p. 3).

There are several aspects to consider here. First, it is claimed that credibility is a probability, and second, that it is expressed without reference to a proposition. Formally, this seems to mean that credibility is a marginal (or unconditional) probability, written $\Pr(E, G)$, where E and G denote evidence and ground respectively. Whether it makes sense to claim that a probability is "assessed without considering the hypothesis in question" (Sunde and Franqueira, 2023, p. 3) is questionable, since in principle a probability such as $\Pr(E)$ can be considered in relation to particular propositions of interest. Using the so-called *extension of the conversation* (e.g. Lindley, 2006) (or the law of total probability), one can write, for example, $\sum_i \Pr(E \mid H_i)\Pr(H_i)$. Thus, conditioning

[3] Here, the term strength of findings can be equated with probative value of evidence (findings).

[4] As an aside, this is yet another example of SF suggesting that the strength evidence could be assessed with respect to a proposition in isolation. See our discussion in Section 3.2 for why this is not sound.

hypotheses are part of a coherent framework of reasoning, unless one actively ignores them.

Moreover, the assertion that credibility is a "non-numerical probability" (Sunde and Franqueira, 2023, p. 3) once again falls short of the technical background of the notion of probability invoked here. Taking the liberty of treating the notion of probability lightly in a non-numerical way does not make the numerical structure of probability disappear. It suffices to consider that the degree of uncertainty of an individual about the truth or falsity of an event of interest, expressed in terms of probability, can in principle be elicited quantitatively, e.g. by comparison with a standard (e.g. Lindley, 1985, 2006; von Winterfeldt and Edwards, 1986). Numbers are therefore an integral part of the concept of probability.

It is deeply problematic for forensic examiners to disregard and silence the quantitative structure of the concept of probability. At the very least, using the terminology without ensuring an awareness and understanding of its underlying properties exposes examiners to the question of whether they understand what they are talking about. This is particularly critical for the concept of probability, as it directly undermines SF's goal of "adding transparency to uncertainty" (Sunde and Franqueira, 2023, p. 1), prominently stated in the title of their paper. Since probability is the scientific measure of uncertainty (Lindley, 2006), being transparent about uncertainty would require one to deal with probability in a sound way, rather than in a half-baked (non-numerical) way.

For the sake of completeness, let us add that SF propose to specify the credibility of "evidence" and "ground" with a so-called C-value, defined as follows: (C1) Very weak, (C2) Weak, (C3) Good, (C4) High, (C5) Very high (Sunde and Franqueira, 2023, p. 4). SF provide explanations for each of these levels of credibility to help examiners make their assignments. The usefulness and practicability of this is not clear to us, especially as some of the descriptions are virtually incomprehensible (e.g. "The meaning is of high elasticity and very inconclusive" (Sunde and Franqueira, 2023, p. 4)).

Let us now turn to SF's treatment of the notion of relevance. SF state that relevance "is assessed by asking: *Assuming that the evidence is true, what is the probability that the hypothesis is true?*" (Sunde and Franqueira, 2023, p. 4, emphasis as in original). As we mentioned in Section 3.3, the italicised phrase technically refers to a posterior probability, written $\Pr(H \mid E)$, where H and E denote the proposition (hypothesis) and the evidence, respectively. However, defining relevance in this way makes no sense, because $\Pr(H \mid E)$ is what evidence evaluation is *ultimately* intended to help with: assessing the probability of a claim in light of the evidence. This makes SF's ABMEO circular, because it seeks to arrive at an evaluative opinion through a process that asks examiners, as part of the process, to make just that ultimate assessment.

This circularity could have been avoided because the concept of relevance is well established in both forensic science and legal literature. In the forensic science literature, Stoney (1991, 1994) introduced relevance as a variable in the development of a likelihood ratio to account for uncertainty about whether the examined item or trace (evidence) came from the offender, i.e., whether the evidence is related to (or a consequence of) the event under investigation. In further structural analyses of this account using Bayesian networks, Garbolino and Taroni (2002) formally showed that relevance is a root variable and clearly distinct from $\Pr(H \mid E)$, contrary to what SF claim.[5] In law, relevance is commonly understood as the tendency of a piece of evidence to make the probandum more or less probable (e.g. Rule 401, FRE). In turn, the *strength* of this inferential support is called probative force, which is to be distinguished from relevance. SF misconceive this fundamental distinction when they write that "[t]he *relevance* of an item of evidence indicates how strongly does it [sic] support the hypothesis in question" (Sunde and Franqueira, 2023, p. 3–4, emphasis added).

A last comment on SF's use of the term "relevance" is necessary for our discussion below. Note that, in analogy to their notion of credibility, SF define relevance using a scale of five R-values: (R1) Very weak, (R2) Weak, (R3) Good, (R4) High, (R5) Very high (Sunde and Franqueira, 2023, p. 4). SF provide a table with verbal explanations to help examiners assign R-values. But again, we do not think this is feasible or recommendable, as it instructs examiners to opine on the ultimate issue, in terms of $\Pr(H \mid E)$, which they – and forensic scientists in general – clearly ought to refrain from doing.

Based on the elements introduced above, we now turn to SF's concept of Argument Matrix. The Argument Matrix is designed for the "assessment of the probative value of an argument" (Sunde and Franqueira, 2023, p. 4). SF present the Argument Matrix in the form of a 5×5 table, where the rows are labelled with R-values and the columns with C-values. Each element of the Argument Matrix represents the minimum of the respective pair of R- and C-value. For example, the cell in row R4 (High) and column C1 (Very weak) contains the assignment "Very weak", the cell in row R2 (Weak) and column C3 (Good) contains the assignment "Weak", and so on. More generally, we can use the notation introduced above to formally state SF's claim about the probative value of an argument. For convenience, let us write PV_A for the probative value of an argument A. SF's claim about the probative value of an argument A can thus be written as follows:

$$\mathrm{PV}_A = \min\{\underbrace{\Pr(E,G)}_{\text{C-value}}, \underbrace{\Pr(H \mid E)}_{\text{R-value}}\}. \tag{1}$$

Recall that SF define credibility and relevance in terms of (qualitatively expressed) probabilities. Equation (1) thus tells us that the probative value of an argument is given by the minimum of two probabilities representing credibility and relevance.

There are several problems with Equation (1). First, as noted in Section 3.1, there is no point in talking about the probative value of an argument, because an argument is not evidence. Second, even if we accept the conjecture that an argument can have probative value, it is not clear how this probative value can be combined *coherently*[6] and without redundancy with the probative value of the evidence (as defined in conventional probabilistic terms), since, as Equation (1) suggests, PV_A is defined as a function of the evidence. Third, according to Equation (1), probative value would be a number between zero and one because SF define C- and R-values as probabilities. However, there is no formal relationship in the form of Equation (1) known in probability theory for probative value. In the probabilistic account, probative value is a factor that quantifies the change in the odds of the propositions of interest on receipt of information (e.g. Aitken et al., 2020). This factor can take values in a much wider range, i.e. 0 and ∞. See also Box 1 for a more formal demonstration of a counterexample of the way in which SF's Argument Matrix suggests that credibility and relevance assessments should be combined.

Overall, the above considerations therefore show that SF's claim that the ABMEO is "founded on [...] key principles and concepts from probability" (Sunde and Franqueira, 2023, p. 9) is groundless. In fact, the ABMEO violates those fundamental principles and therefore does not comply with the ENFSI Guideline (Willis et al., 2015). As such, SF's proposal of the ABMEO is an example of the dangers of holistic approaches that can lead to (digital evidence) exceptionalism (Biedermann and Kotsoglou, 2020). In the context here, digital evidence exceptionalism manifests itself in the idea that for reasoning under uncertainty one can simply make up seemingly plausible rules without first building a logically sound underlying structure. It is therefore worth recalling Lindley's old warning about the importance of coherence, which unfortunately receives too little attention today in the forensic science literature (Lindley, 1985, p. 37, emphasis added), especially in DFS:

[5] See also Box 1 for an example of the use of this definition of relevance.

[6] Here, coherence is taken to mean agreement with the rules of probability.

Credibility and relevance in the probabilistic account of probative value: a counter-example to SF's Argument Matrix

Here we provide a brief example to illustrate two points: first, that the notions of credibility and relevance do *not* combine in the way suggested by SF's Argument Matrix, and second, that a categorical rule-based system, as suggested by SF's Argument Matrix, is redundant. We begin by articulating our basic assumptions and explaining the differences between our conceptual (probabilistic) framework and SF's non-numerical account (ABMEO, including the Argument Matrix):

- We focus on characterising the probative value of the *evidence*, rather than an argument as in SF's account, for the reasons explained in Section 3.1.
- We use the likelihood ratio as the measure of the probative value of the evidence (Aitken et al., 2020) as recommended by the ENFSI Guideline (Willis et al., 2015). We denote the value of the evidence by V.
- Following earlier accounts in the forensic science literature (Evett et al., 1998; Stoney, 1991, 1994), we define relevance as the probability that the evidential material (of unknown source) was left by the perpetrator. As noted in Section 3.4, this is slightly different from, but not incompatible with, accounts in the legal literature that define relevance as the tendency of a piece of evidence to make a probandum more or less probable.
- We define credibility, following Taroni et al. (2021), as an attribute of the source of information (e.g., testimony, a scientist's report, etc.) regarding an event of interest (e.g. analytical features of examined trace material).

The hypothetical case we consider here, loosely adapted from Willis et al. (2015, p. 12), involves a single trace found at the scene of a crime committed by a single perpetrator (see Evett (1993) for an extension to k perpetrators). We assume that the examiner of the trace is perfectly credible, i.e. his/her testimony is unequivocal. The finding, evidence E, is that the stain is human blood. We also assume that the trace is perfectly relevant: it is fresh, present in large quantities and found on a window that was broken as a result of the crime, as confirmed by video surveillance. Denote this conditioning information by I. Suppose further that a person of interest (POI) is available. The propositions of interest are "The POI is the offender" (H_p) and "An unknown person is the offender" (H_p).

The likelihood ratio $\mathrm{V} = \Pr(E \mid H_p, I)/\Pr(E \mid H_d, I)$ for the above case example can be shown to correspond to (Taroni et al., 2014):

$$\mathrm{V} = \frac{r + (1-r)\gamma}{r\gamma + (1-r)[p + (1-p)\gamma]} \tag{2}$$

where p is the probability that the POI left the trace for innocent reasons, r is the probability of relevance, and γ is the proportion of the population of interest that has the analytical characteristics of the trace.

Suppose the POI had no opportunity to leave a trace for innocent reasons ($p = 0$) and that the trace is relevant ($r = 1$). Finally, set $\gamma = 1$ because all members of the population have human blood. Obviously, for such a situation, the likelihood ratio, Equation (2), is 1.

One might object that this example is absurd, because the forensic finding that the fluid collected at the crime scene is human blood is obviously not helpful in distinguishing between humans, since all humans have human blood in their bodies. But that is precisely our point, and the likelihood ratio correctly translates this intuition by a value of 1. The key point we want to make here is that one can have a forensic report from a perfectly credible source of information about a trace that is perfectly relevant, and yet the evidence has *no* probative value with respect to the question of who is the offender (or even the source of the stain). So it is simply not the case, as SF's Argument Matrix suggests, that high credibility and high relevance imply high probative value. This observation illustrates the drawbacks of holistic and non-symbolic approaches that attempt to devise general rules for assessing probative value.

We can even go one step further and make the additional point that there is *no* fixed relationship between relevance and credibility, contrary to what SF's Argument Matrix suggests. Suffice it to consider a variation of the above example by assuming that the analytical features of the trace are to some extent discriminative. To illustrate this point, suppose we focus on the blood group for which the examiner reports that it is type A (and the POI also has type A). Let γ for type A in the relevant population be 0.45. Keeping the other assignments as above, we can now see that the likelihood ratio (for both the source and crime level propositions) is $1/\gamma$, i.e. about 2. Next, suppose we focus on a forensic DNA profile for which γ is 1 in 1 billion. The likelihood ratio would then be 1 billion. In all of these alternative case examples, however, the assessment of credibility and relevance remains unchanged.

One might also object that SF deal with digital evidence, while our example deals with physical evidence. However, this observation distracts from the main problem. SF claim to provide a *general* account of how to assess probative value, as a supplement to the ENFSI Guideline (Willis et al., 2015). In fact, SF write: "It [the ABMEO] is flexible and may be applied for targeted analysis of one or a few pieces of evidence from a single forensic domain, such as digital forensics, as demonstrated in this paper. However, it may also be feasible for a holistic analysis of *different types of evidence* from *different domains*" (Sunde and Franqueira, 2023, p. 2, *emphasis added*). They also repeat this claim in their discussion (Sunde and Franqueira, 2023, p. 8). In addition, SF discuss examples involving different types of evidence (e.g. information from a public register, eyewitness evidence and a forensic examiner's "identification" based on facial images). The ABMEO should therefore stand the test of coherent handling of evidence, *regardless* of its nature. We show here why it does not, even in the simple(st) and favourable case of a single trace and a single offender. It is therefore unclear how the ABMEO could hope to deal with more complex examples.

In summary, we can see that the likelihood ratio provides a self-contained framework for quantifying the probative value. Notwithstanding this, we acknowledge and agree with Professor Kaye's point that "[...] no mathematical result is self-applying, and additional argument is necessary to bridge the gap from a general mathematical truth to a substantive application" (Kaye, 1999, p. 27). Here, bridging the gap consists of substantiating and justifying how the different components of the likelihood ratio relate to the case of interest. However, this does not require a separate theory of argumentation as proposed by SF (Sunde and Franqueira, 2023). The ABMEO is therefore redundant.

Box 1. A probabilistic counter-example to SF's Argument Matrix (Evett, 1993; Evett et al., 1998; Kaye, 1999; Taroni et al., 2014, 2021).

"The key point [...] is that the ways in which statements of uncertainty can be combined are by no means arbitrary. *One cannot sit down and think up apparently reasonable rules*. The only ones are the three we have discussed: no more, no less. Some workers have suggested other types of combination [...]. These ideas are false, because one is not free to engage in the intellectual exercise of law creation. The laws are forced upon you. It is a case of the inevitability of probability. The laws ensure that several statements of uncertainty cohere."

3.5. On normative approaches in evidence evaluation

In the opening sentence of Section 2 of SF's paper, the authors claim, citing Verheij (2012) in support, that "[t]here are three *normative* approaches to reasoning with legal evidence: probabilistic, narrative or

scenario and argumentation-based approaches" (Sunde and Franqueira, 2023, p. 2, emphasis added). This statement is incorrect in two respects. First, it is incorrect in that it misquotes Verheij (2012), who only mentions "three *types* of approaches" (Verheij, 2012, p. 174, emphasis added). He does not use the term "normative" throughout his paper. Secondly, the statement is misleading in the sense that it seems to suggest that narrative or scenario- and argumentation-based approaches are normative in the same way as the probabilistic approach. However, this is not the case.

A probabilistic approach is normative in the sense that it provides a standard or norm for measuring uncertainty (i.e. using probabilities) and for combining measures of uncertainty (i.e. the rules of probability theory) to ensure coherence (Lindley, 1982, 1985). However, this is not an anything-goes approach. In normative accounts, a standard must be justified *independently* of actual human reasoning and decision-making behaviour (Baron, 2012), a point often misunderstood in the forensic science literature (Biedermann et al., 2014). It is unclear how narrative or scenario- and argumentation-based approaches could be normative in this sense without actually incorporating probability as the core method for reasoning under uncertainty, which would render the initial distinction between the approaches meaningless. Instead, narrative or scenario- and argumentation-based approaches might be better classified as purely *descriptive* approaches. This type of approach, as the name suggests, provides a descriptive account of how people reason in real-world circumstances. Indeed, constructing narratives, scenarios and arguments is what people do when they try to make sense of evidence. However, in such descriptive approaches there is no guarantee that the process or its outcome is coherent, let alone accurate, precisely because of a lack of a standard. Descriptive approaches are therefore of limited use unless their output is checked against a standard. This understanding has been recognised and well documented in the statistical and applied psychology literature on judgement and decision making for decades (e.g. Baron, 2008; de Finetti, 1961).

SF's ABMEO inadvertently provides an example of the drawbacks that can result from not properly distinguishing between the normative perspective on the one hand and the descriptive perspective on the other. SF claim that the ABMEO incorporates elements of probability, thus giving the impression that it has a normative status. This is misleading, however, because the ABMEO defines its own rule-based procedure (i.e. the Argument Matrix), which deliberately overrides and sweeps away any properly normative input, such as a likelihood ratio. The ABMEO cannot therefore claim normative status, especially in view of the numerous problems and inconsistencies we have outlined in Sections 3.1 to 3.4 (and Box 1). Overall, then, the ABMEO has things exactly backwards: instead of developing an argument, where necessary and useful, to support probabilistic analysis, the ABMEO places the malleable notions of "narrative," "scenario," and "argument" at the forefront in a holistic way.

Interestingly, the idea that one can freely invent theories for evidence evaluation is not new to the DFS literature. Digital Evidence Certainty Descriptors (DECDs) (Horsman, 2020) are a previously proposed idea (see Biedermann and Kotsoglou (2020) for a critique) that suffers from the same kind of drawbacks we have highlighted here for SF's ABMEO. Both accounts amount to what we have elsewhere called digital evidence exceptionalism. The proliferation of such widely divergent metatheories, most of which are verbose accounts of intuitive modes of reasoning, unfortunately without providing a coherent underlying conceptual architecture, is a cause for concern and a distraction from the real problem: the assessment of the value of *evidence* using a scientific measure of uncertainty.

3.6. On probability

The vagueness surrounding SF's use of the term "normative," as revealed in the previous section, is also reflected in their account of probability for evidence evaluation. Here, SF invoke three general and stereotypical criticisms that warrant several comments.

3.6.1. Availability of data

In Section 2 of their paper, SF state that "[t]he main critique [of the probabilistic approach] is that a knowledge base from which calculations may be obtained rarely exists in digital forensics" (Sunde and Franqueira, 2023, p. 2). However, this misses the point entirely. The lack of data is primarily an aspect of the situation at hand, not of the method used to deal with it.

The value of the probabilistic perspective is, first, to help define the relevant question(s), i.e. the questions that the scientist can and cannot legitimately help to answer. In the context of forensic evidence evaluation, the focus is on "the probability of the evidence given the propositions" (Evett et al., 2000, p. 234). As an aside, we should also recall Glenn Shafer's statement in a personal communication with Judea Pearl: "probability is not really about numbers; it is about the structure of reasoning" (Pearl, 1988, at p. 15).

Secondly, it is precisely one of the strengths of probability theory to reveal the extent of the available data and knowledge. For example, if there is no, or not enough, relevant data and knowledge available, and therefore the examiner is unable to articulate probabilities (and hence a likelihood ratio), then the evidence is, and should be, considered *uninterpretable* in the situation at hand (Biedermann and Kotsoglou, 2022). This in itself is a pertinent finding and not a failure of the method. For reasons of transparency, uninterpretability should be brought to the attention of the recipients of expert information. On the other hand, a situation of uninterpretability should not be seen as a reason for forensic examiners to treat the evidence with an alternative inferential method of unknown credentials, as this would amount to – in Salmon's words – an attempt to perform "epistemological magic" (Salmon, 1966, p. 66). As he points out, "[...] there are ways of transforming ignorance into knowledge–by further investigation and the accumulation of more information. It is the same with all "magic": to get the rabbit out of the hat you first have to put him in" (Salmon, 1966, p. 66).

3.6.2. Atomism and dependencies

SF claim that the use of probability to evaluate evidence is limited in the sense that it amounts to an "atomic approach" which "does not consider dependencies between items of evidence" (Sunde and Franqueira, 2023, p. 2). This is a rather bizarre reproach that falls short of the actual properties of probability theory and auxiliary concepts, such as graphical models, that help in its practical implementation.

To begin with, it is simply not true that probability (theory) does not allow one to consider "dependencies between items of evidence" (Sunde and Franqueira, 2023, p. 2). The laws of probability allow one to deal with dependent events, where "dependent" means that knowledge of the occurrence of one event affects our probability of the occurrence of another event, and vice versa (Aitken et al., 2020, p. 82–90). In addition, for more than 40 years, graphical probabilistic models (Bayesian networks) have been available (Pearl, 1982, 1988), the framework *par excellence* for dealing with probabilistic relevance relationships.

Next, the claim that probability (theory) implies an atomic approach is primarily a consequence of the way in which probability is sometimes used, rather than a property or limitation of probability itself. Of course, when one considers Bayes' theorem for a single item of evidence E and a single pair of propositions H, and perhaps does so repeatedly for different items of evidence, one is indeed *choosing* an atomic perspective. But, this way of proceeding is not imposed. While Bayes' theorem can indeed be applied iteratively, regardless of the order in which the items of evidence are considered, the theorem can also be applied in a single step, treating the multiple items of evidence in combination. This does not mean that *in practice* there are no limits to the number of variables that can be included in an analysis and to the density of the dependency structure among those variables, but these are limits on the part of

the user, not on parts of the theory.[7] Put another way, complexity arises from the world around us, it is not the fault of probability theory (Friedman, 1992). As Prof. Friedman puts it: "[...] any theory that could not in principle represent the complexity surrounding us would have limited value" (Friedman, 1997, p. 288). Note that we make no claim here about the suitability of probability (theory) for evidence evaluation in the legal process in general, which is a separate topic. We are only examining the criticisms that SF make of probability on the theoretical side.

3.6.3. Completeness of evidence

The third criticism that SF advance against probability theory concerns the completeness of evidence. They express their criticism with the following single sentence, referring to a monograph by Tecuci et al. (2016), without any further discussion or justification: "Another limitation pointed out is that it [probability] does not consider the completeness of evidence" (Sunde and Franqueira, 2023, p. 2). This leaves the reader wondering what to make of this statement, not least because its suitability as a critique of probability is doubtful.

The alleged problem of evidence completeness refers to the observation that a probability assigned on the basis of an *incomplete* collection of items of evidence at time t_1 may be overturned or reversed, in either direction, at a time t_2, after the receipt of additional evidence not available at time t_1. Arguing in this way alludes to the frequentist statistical preoccupation with other data that might be observed, but that are not available. The concern is that data could have been encountered *other* than those actually observed. But this is a truism. Lindley (1985) has called this hindsight, i.e. judging an ex ante situation from an ex post perspective. Understood in this way, however, evidence completeness cannot serve as a criticism of probability. The purpose of probability is to ensure coherence in organising one's degrees of belief in the light of the evidence available *at the time an assessment is to be made*. If one receives additional evidence later, the laws of probability can again be invoked to revise one's beliefs – in a rational way. However, the result will not be a correction of the former probability or an indication that the previous probability was somehow "wrong", because the two probability assignments at t_1 and t_2 are conditioned by different sets of evidence. Over time we accumulate evidence and our knowledge (base) becomes richer, so it is natural that probability assignments evolve as well (de Finetti, 2017). In this sense, evidence (in)completeness is merely an inevitable aspect of the real world. Rather than demonstrating a problem with probability, evidence (in)completeness demonstrates the necessity of probability. If we want to argue that a newly received item of information ought to change our current belief, then coherence requires that we process that item of evidence according to the rules of probability.

3.7. On bias reduction

SF repeatedly suggest that their method, through its focus on transparency and the obligation to assess findings against competing propositions, might help reduce cognitive bias (e.g., "[t]he process of articulating argument components may counter the confirmation bias" (Sunde and Franqueira, 2023, p. 9)), or at least "make a possible bias more visible to peers, which increases the possibility of bias detection and correction" (Sunde and Franqueira, 2023, p. 9). However, these are largely unsubstantiated claims. In particular, if the claim is that mere insistence on and awareness of transparency, i.e. willpower alone, could mitigate bias, then the suggestion is prone to fall into a common misconception about cognitive bias, called the illusion of control (Dror, 2020). Put differently, mere awareness or visibility of bias is not a sufficient means to combat bias. Rather, once one has become aware of a possible bias, or it has become visible, specific *additional* measures or actions, such as avoiding exposure to (additional) task-irrelevant information, are needed to deal with the identified problem.

Overall, it remains unclear what exactly SF claim in terms of bias reduction. On the one hand, they mention that the ABMEO has the *potential* to reduce bias, while on the other hand they instantly tamper this claim by asserting that "the proposed method involves subjectivity and discretion, it does not entirely eliminate cognitive bias influencing the many judgements and decisions made throughout the process" (Sunde and Franqueira, 2023, p. 9).

3.8. Legal requirements for reliable opinions, the problem of measurement and the standard of proof

The law in the jurisdiction where one of the authors (SF) works (and where one of us works), namely England and Wales (similar arguments can be made about any other jurisdiction), renders any opinion evidence inadmissible unless that evidence has a sufficiently reliable scientific basis (*R v Dlugosz* [2013] EWCA Crim 2, see also Kotsoglou and Biedermann (2022)). This is the essence of the *opinion rule*. Fact-finders, in particular the jury, do not need, and indeed are not allowed, to hear anyone's opinion on anything which witnesses have not seen or heard for themselves, unless the latter have a reliable basis to back their claims. For the jury can form its own opinion on matters of common experience, i.e. matters that do not go beyond laypeople's knowledge and understanding of the world. If an opinion is not *reliable* (i.e. not based on a valid methodology) or *helpful* (*R v Turner* [1975] 1 All ER 70), then it would be *inadmissible* as evidence.

Thus, the question now is: Do SF provide such a reliable method? As we have shown throughout this article, the answer is: No – by no stretch of the imagination or the meaning of the word "reliable". At its heart, the ABMEO faces the insurmountable problem of *measurement* and uncertainty quantification. While probability theory in principle addresses the problem of measuring uncertainty scientifically, we have shown that the ABMEO violates probabilistic principles in several ways (Section 3.4).

Measurement is a ubiquitous aspect of our modern world (Baker and Hacker, 2005). However, in many practical situations, such as the application of science in the legal process, measurement cannot be an arbitrary activity where we make up the rules as we go along. We need canonical samples as points of reference, as we do in other contexts, for example when measuring length in relation to the standard set by the Standard Metre Bar. It is unclear what the canonical sample would be in the context of the ABMEO. Would it perhaps be a "holistic" *feeling*, i.e. a person's sensation that the credibility of evidence is "weak" or "high", or even "good", irritatingly using axiological/moral terms such as "good" in the very process of measurement? This would be reminiscent of the apocryphal diviner who states: "I feel in my hand that the water is three feet under the ground", a statement to which the British-Austrian philosopher Wittgenstein says "we should like to answer: ‹I don't know what this *means*›" (Wittgenstein, 1958, Blue Book). To hear someone say that he or she *feels* "this pencil is five inches long" (Wittgenstein, 1958, Blue Book) makes sense only because of the practice of measurement, which can verify the rectitude of this assessment. We simply have to use a ruler to find out. Unfortunately, there is no ruler or measuring process for SF's descriptors such as "weak" or "high".[8] For this reason, SF's component descriptors for the "probative value of an argument" (Sunde

[7] Again, as mentioned above, the advent of graphical implementations of probabilistic reasoning systems has pushed the boundaries of these limitations. For some critical and demanding commentators, these boundaries may still be far from making the framework usable in practice, but any fair assessment of the theoretical development must conclude that it is far from *imposing* an atomic approach.

[8] Although SF provide descriptions of how to assign each of their proposed five C- and R-values, such an informal approach is notoriously unreliable in the sense that it leads to noisy output when applied by different persons.

and Franqueira, 2023, p. 3) are strictly meaningless in a formalised context such as the criminal process.

SF claim that their method "[...] allows the legal decision-makers [sic] insight into essential aspects when considering whether the evidence is sufficient to meet the standard often expressed as «proven beyond any reasonable doubt»" (Sunde and Franqueira, 2023, p. 8–9). In light of our discussion above, we find this claim unsupported. Furthermore, as regards the jurisdiction to which SF frequently refer, i.e. England and Wales, their claim is misrepresenting the law. In England and Wales, the standard of proof is feeling "sure" [Crown Court Compendium, at 5-2], not "proven beyond *any* reasonable doubt" (Sunde and Franqueira, 2023, p. 9, emphasis added).

4. Discussion and conclusions

Throughout their paper, SF repeatedly emphasise that their proposed ABMEO amounts to a *holistic* approach that may "be used as a stand-alone method or in combination with other methods." (Sunde and Franqueira, 2023, p. 2). It seeks to formalise a broader framework for reasoning, of which hard-wired probabilistic value of evidence assessments can be, but are not necessarily required to be, a part. They state: "As individual pieces of evidence may be subject to Bayesian probabilistic analyses, the proposed argument-based method can be applied as an "add-on" [...]. It is flexible and enables the combination of multiple [sic] evidence of similar kinds as well as different types of evidence." (Sunde and Franqueira, 2023, p. 8) The authors go on to claim that their method is at the level of Kahneman's "system 2 thinking" because of the emphasis on "[...] explicating the components and logic within each argument" (Sunde and Franqueira, 2023, p. 9).

We find these claims unfounded. To begin with, the assertion that the result of a probabilistic analysis, such as a likelihood ratio, can be used as part of another reasoning framework is a truism, but the problem lies elsewhere. SF's ABMEO contains no formalism to ensure that a probabilistic result *coherently* informs subsequent reasoning steps, where *coherence* is understood in the strict sense of adherence to the logic of reasoning under uncertainty (i.e. probability theory). For example, in their first case example, they mention a likelihood ratio of 5 (Sunde and Franqueira, 2023, p. 6). This result completely dissolves in their subsequent processing of this likelihood ratio when using the ABMEO. Proceeding in this way is not compliant with the ENFSI Guideline (that SF claim to follow), which recommends reporting the assigned value of the evidence (i.e. the likelihood ratio). It is also not clear what the "logic" is that SF claim their method "explicates" (Sunde and Franqueira, 2023, p. 9) because, as we have shown in Section 3.4 (and Box 1), SF's argument matrix finds no justification when analysed in probabilistic terms.

In light of these observations, we find SF's suggestions that "[s]uch support scales allow practitioners *without competency* and experience in quantitative approaches to use the proposed method" (Sunde and Franqueira, 2023, p. 9, emphasis added) and that the ABMEO "may also be feasible for a holistic analysis of different types of evidence from different forensic domains" (Sunde and Franqueira, 2023, p. 2) highly problematic, if not downright dangerous. Under no circumstances should non-competent individuals be tasked with producing evaluative opinions, in digital forensic science or in any other branch of forensic science. To suggest otherwise is a perfect example of (digital) forensic exceptionalism (Biedermann and Kotsoglou, 2020). If SF disagree, they should ask themselves whether they would find it comforting to be defendants in a case where the forensic examiner charged with providing an evaluative opinion lacks both scientific competence and experience.

We also take issue with the suggestion that the ABMEO could somehow complement conventional (i.e. probabilistic) assessments of the value of evidence, particularly in situations where the application of the latter faces challenges at the level of probability assignment. This would suggest that forensic examiners could magically produce an evaluative opinion *despite* being unable to ascertain the underlying probabilistic structure. This is wishful thinking. Whenever forensic examiners cannot assign a value according to the recommendations in the ENFSI Guideline, they should say so. This would be a case of evidence uninterpretability (Biedermann and Kotsoglou, 2022), which in itself is a result that should be reported as such, i.e. that the evidence provides *no* assistance to any recipient of expert information who is asked to judge which of the competing propositions is true. To deviate instead, as SF's ABMEO suggests, to a "holistic analysis" (Sunde and Franqueira, 2023, p. 2) in which vague terms such as "weak", "good" or "high" (this enumeration is not exhaustive) are combined in a way (i.e. SF's Argument Matrix) that finds no justification in a coherent reasoning framework, is to turn a blind eye to the methodological, socio-legal and doctrinal requirements of scientific evidence in the first place.

In conclusion, we find that on all the points we have analysed, SF's account is not only incorrect, but deeply flawed and openly dangerous. SF write that "[t]he proposed method [...] leads to the elaboration of an evaluative expert opinion which follows logically from the assessment of the opposing hypotheses." (Sunde and Franqueira, 2023, p. 9) As we have repeatedly pointed out, forensic examiners simply have no business in "the assessment of the opposing hypotheses" (Sunde and Franqueira, 2023, p. 9). Instead, they should assess (the value of) their findings. Moreover, in practice, an evaluative opinion does not *logically follow* from (i.e. as consequence of) the assessment of opposing hypotheses.[9] It is exactly the other way round: the assessment of opposing hypotheses is *post hoc* to the receipt of an evaluative opinion from a forensic examiner. In addition to the numerous terminological confusions (e.g. regarding the notion of probative value and normativism), the transposed conditionals and the stereotypical criticism of probability concepts, SF's account is unfortunately an example of the fact that digital forensic science is still largely struggling on its way to coherent evaluative principles in the sense emphasised by current guidance documents and recommendations (e.g. Aitken et al., 2010; Nic Daéid et al., 2020; Willis et al., 2015; The Council of the Inns of Court (COIC) and the Royal Statistical Society (RSS), 2017).

CRediT authorship contribution statement

Alex Biedermann: Writing – review & editing, Writing – original draft, Conceptualization. **Kyriakos N. Kotsoglou:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that this paper was written in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Acknowledgements

The authors are grateful to Dr Itiel Dror for comments on cognitive bias.

Data availability

No data was used for the research described in the article.

References

Aitken, C.G.G., Roberts, P., Jackson, G., 2010. Fundamentals of Probability and Statistical Evidence in Criminal Proceedings (Practitioner Guide No. 1), Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Royal Statistical Society's Working Group on Statistics and the Law.

[9] We acknowledge, however, that it is possible to use Bayes' theorem in reverse mode (e.g. Good, 1983), but in practice this is incompatible with the way in which forensic evidence is used in legal proceedings.

Aitken, C.G.G., Taroni, F., Bozza, S., 2020. Statistics and the Evaluation of Evidence for Forensic Scientists, third ed. John Wiley & Sons, Chichester.

Baker, G.P., Hacker, P.M.S., 2005. Wittgenstein: Understanding and Meaning, 2nd ed. An Analytical Commentary on the Philosophical Investigations, vol. 1. Blackwell Publishing, Malden.

Baron, J., 2008. Thinking and Deciding, 4th ed. Cambridge University Press, New York.

Baron, J., 2012. The point of normative models in judgment and decision making. Front. Psychol. 3, 577, 1–3.

Biedermann, A., Kotsoglou, K., 2020. Digital evidence exceptionalism? A review and discussion of conceptual hurdles in digital evidence transformation. Forensic Sci. Int.: Synergy 2, 262–274.

Biedermann, A., Kotsoglou, K., 2022. (Un-)interpretability in expert evidence: an inquiry into the frontiers of evidential assessment. Quaestio facti (Rev. Int. sobre Razonamiento Probatorio Quaestio facti. Int. J. Evidential Legal Reason.) 3, 481–515.

Biedermann, A., Taroni, F., Aitken, C., 2014. Liberties and constraints of the normative approach to evaluation and decision in forensic science: a discussion towards overcoming some common misconceptions. Law Probab. Risk 13, 181–191.

Champod, C., Lennard, C., Margot, P., Stoilovic, M., 2016. Fingerprints and Other Ridge Skin Impressions, second ed. CRC Press, Boca Raton.

de Finetti, B., 1961. Dans quel sens la théorie de la décision est-elle et doit-elle être “normative”. In: Colloques Internationaux du Centre National de la Recherche Scientifique, “La Décision”. Paris, pp. 159–169. Éditions du Centre National de la Recherche Scientifique.

de Finetti, B., 2017. Theory of Probability, A Critical Introductory Treatment, reprint ed. John Wiley & Sons, Chichester.

Dror, I.E., 2020. Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. Anal. Chem. 92, 7998–8004.

Edmond, G., Martire, K., 2018. Antipodean forensics: a comment on ANZFSS’s response to PCAST. Aust. J. Forensic Sci. 50, 140–151.

Evett, I.W., 1993. Establishing the evidential value of a small quantity of material found at a crime scene. J. Forensic Sci. Soc. 33, 83–86.

Evett, I.W., 2009. Evaluation and professionalism. Sci. Justice 49, 159–160.

Evett, I.W., 2015. The logical foundations of forensic science: towards reliable knowledge. Philos. Trans. R. Soc. Lond. B, Biol. Sci. 370, 1–10.

Evett, I.W., Jackson, G., Lambert, J.A., McCrossan, S., 2000. The impact of the principles of evidence interpretation and the structure and content of statements. Sci. Justice 40, 233–239.

Evett, I.W., Lambert, J.A., Buckleton, J.S., 1998. A Bayesian approach to interpreting footwear marks in forensic casework. Sci. Justice 38, 241–247.

Evett, I.W., Weir, B.S., 1998. Interpreting DNA Evidence. Sinauer Associates Inc., Sunderland.

Friedman, R.D., 1992. Infinite strands, infinitesimally small: storytelling, Bayesianism, hearsay and other evidence. Cardozo Law Rev. 14, 79–102.

Friedman, R.D., 1997. Answering the Bayesioskeptical challenge. Int. J. Evid. Proof, Spec. Issue 1, 276–291.

Garbolino, P., Taroni, F., 2002. Evaluation of scientific evidence using Bayesian networks. Forensic Sci. Int. 125, 149–155.

Gill, P., Hicks, T., Butler, J.M., Connolly, E., Gusmão, L., Kokshoorn, B., Morling, N., van Oorschot, R., Parson, W., Prinz, M., Schneider, P.M., Sijen, T., Taylor, D., 2018. DNA commission of the International society for forensic genetics: Assessing the value of forensic biological evidence – Guidelines highlighting the importance of propositions. Part I: evaluation of DNA profiling comparisons given (sub-) source propositions. Forensic Sci. Int.: Genet. 36, 189–202.

Good, I.J., 1983. Good Thinking: the Foundations of Probability and Its Applications. University of Minnesota Press, Minneapolis.

Hicks, T., Buckleton, J., Castella, V., Evett, I., Jackson, G., 2022. A logical framework for forensic DNA interpretation. Genes 13, 1–26.

Horsman, G., 2020. Digital evidence certainty descriptors (DECDs). Forensic Sci. Int. Digit. Investig. 32, 200896.

Kaye, D.H., 1999. Clarifying the burden of persuasion: what Bayesian decision rules do and do not do. Int. J. Evid. Proof 3, 1–29.

Kotsoglou, K.N., Biedermann, A., 2022. Inroads into the ultimate issue rule? Structural elements of communication between experts and fact finders. J. Crim. Law 86, 223–240.

Lindley, D.V., 1982. Scoring rules and the inevitability of probability. Int. Stat. Rev. (Rev. Int. Stat.) 50, 1–11.

Lindley, D.V., 1985. Making Decisions, second ed. John Wiley & Sons, Chichester.

Lindley, D.V., 2006. Understanding Uncertainty. John Wiley & Sons, Hoboken.

Murphy, E., 2009. What ‘Strengthening Forensic Science’ today means for tomorrow: DNA exceptionalism and the 2009 NAS Report. Law Probab. Risk 9, 7–24.

Nic Daéid, N., Biedermann, A., Champod, C., Hutton, J., Jackson, G., Kitchin, D., Neocleous, T., Spiegelhalter, D., Willis, S., Wilson, A., 2020. The Use of Statistics in Legal Proceedings: A Primer for Courts. The Royal Society and The Royal Society of Edinburgh, Edinburgh.

Pearl, J., 1982. Reverend Bayes on inference engines: a distributed hierarchical approach. In: Proceedings of the National Conference on Artificial Intelligence (AAAI’82). Pittsburgh, 1982, August 18–20. Morgan Kaufmann, pp. 133–136.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers, Inc., San Mateo.

Robertson, B., Vignaux, G.A., Berger, C.E.H., 2016. Interpreting Evidence. Evaluating Forensic Science in the Courtroom, second ed. John Wiley & Sons, Chichester.

Saks, M.J., 2007. Remediating forensic science. Jurimetrics 48, 119–124.

Salmon, W.C., 1966. The Foundations of Scientific Inference. University of Pittsburgh Press, Pittsburgh, PA.

Stoney, D.A., 1991. Transfer evidence. In: Aitken, C.G.G., Stoney, D.A. (Eds.), The Use of Statistics in Forensic Science. Ellis Horwood, New York, pp. 107–138.

Stoney, D.A., 1994. Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. J. Forensic Sci. Soc. 34, 17–21.

Sunde, N., Franqueira, V.N.L., 2023. Adding transparency to uncertainty: an argument-based method for evaluative opinions. Forensic Sci. Int. Digit. Investig. 47, 301657.

Swofford, H.J., Cole, S.A., King, V., 2021. Mt. Everest – we are going to lose many: a survey of fingerprint examiners’ attitudes towards probabilistic reporting. Law Probab. Risk 19, 255–291.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., Aitken, C.G.G., 2014. Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science, second ed. Statistics in Practice. John Wiley & Sons, Chichester.

Taroni, F., Garbolino, P., Bozza, S., Aitken, C., 2021. The Bayes’ factor: the coherent measure for hypothesis confirmation. Law Probab. Risk 20, 15–36.

Tecuci, G., Schum, D., Marcu, D., Boicu, M., 2016. Intelligence Analysis as Discovery of Evidence, Hypotheses, and Arguments: Connecting the Dots. Cambridge University Press, New York.

The Council of the Inns of Court (COIC), the Royal Statistical Society (RSS), 2017. Statistics and Probability for Advocates: Understanding the Use of Statistical Evidence in Courts and Tribunals. London.

Verheij, B., 2012. Integrating argumentation, narrative and probability in legal evidence (position paper). In: Finlayson, M.A. (Ed.), The Third Workshop on Computational Models of Narrative (CMN’ 2012), pp. 174–175.

von Winterfeldt, D., Edwards, W., 1986. Decision Analysis and Behavioral Research. Cambridge University Press, Cambridge.

Willis, S., McKenna, L., McDermott, S., O’Donell, G., Barrett, A., Rasmusson, B., Nordgaard, A., Berger, C., Sjerps, M., Lucena-Molina, J., Zadora, G., Aitken, C., Lovelock, T., Lunt, L., Champod, C., Biedermann, A., Hicks, T., Taroni, F., 2015. ENFSI Guideline for Evaluative Reporting in Forensic Science, Strengthening the Evaluation of Forensic Results Across Europe (STEOFRAE). Dublin.

Wittgenstein, L., 1958. The Blue and Brown Books. Harper and Row, New York.